

University of Dundee

Neural correlates of social exchanges during the Prisoner's Dilemma game in depression

Gradin, V. B.; Pérez, A.; Macfarlane, J. A.; Cavin, I.; Waiter, G.; Tone, E. B.

Published in:
Psychological Medicine

DOI:
[10.1017/S0033291715002834](https://doi.org/10.1017/S0033291715002834)

Publication date:
2016

Document Version
Peer reviewed version

[Link to publication in Discovery Research Portal](#)

Citation for published version (APA):

Gradin, V. B., Pérez, A., Macfarlane, J. A., Cavin, I., Waiter, G., Tone, E. B., Dritschel, B., Maiche, A., & Steele, J. D. (2016). Neural correlates of social exchanges during the Prisoner's Dilemma game in depression. *Psychological Medicine*, 46(6), 1289-1300. <https://doi.org/10.1017/S0033291715002834>

General rights

Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from Discovery Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Psychological Medicine

Neural correlates of social exchanges during the Prisoner's Dilemma game in depression --Manuscript Draft--

Manuscript Number:	PSM-D-15-00998R1
Full Title:	Neural correlates of social exchanges during the Prisoner's Dilemma game in depression
Article Type:	Original Article
Corresponding Author:	Victoria Gradin Universidad de la República Montevideo, URUGUAY
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Universidad de la República
Corresponding Author's Secondary Institution:	
First Author:	Victoria Gradin
First Author Secondary Information:	
Order of Authors:	Victoria Gradin Alfonso Pérez Jennifer Macfarlane Ian Cavin Gordon Waiter Erin Tone Bárbara Dritschel Alejandro Maiche Douglas Steele
Order of Authors Secondary Information:	
Manuscript Region of Origin:	UNITED KINGDOM
Abstract:	<p>Background. Depression is a disabling disorder that significantly impacts on the interpersonal functioning of individuals. However, little is known about the neural substrates of such difficulties. In the last few years neuroeconomics, which combines imaging with multiplayer behavioural economic paradigms, has been used to study the neural substrates of normal and abnormal interpersonal interactions.</p> <p>Methods. This study used functional Magnetic Resonance Imaging (fMRI) to investigate neural activity in unmedicated depressed participants (n=25) and matched healthy controls (n=25). During scanning, participants played a behavioural economic game, the Prisoner's Dilemma. In this game, the participant and a co-player independently choose either to cooperate or not cooperate with each other.</p> <p>Results. Depressed participants reported higher levels of negative feelings (betrayal, guilt) during the game than did controls. Neural activation was compared between 'imbalanced' events (when one of the players cooperated and the other defected ('CD' and 'DC')) and 'draw' events (when both players either cooperated or defected ('CC' and 'DD')). Participants preferentially activated the anterior insula and the dorsolateral prefrontal cortex (DLPFC), a region implicated in cognitive control and regulation of emotions. Importantly, compared to controls depressed participants showed reduced activation in the left DLPFC, with the extent of signal reduction correlating with increased self-report feelings of guilt associated with DC outcomes.</p> <p>Conclusions. Our findings suggest that depression is associated with reduced</p>

	activation of the DLPFC during social events that involve unreciprocated cooperation. This abnormality may underlie anomalies in cognitive control and top down regulation of emotions during challenging social exchanges.
--	---

Neural correlates of social exchanges during the Prisoner's Dilemma game in depression

Victoria B. Gradin¹; Alfonso Pérez¹; Jennifer A. Macfarlane²; Ian Cavin²; Gordon Waiter³; Erin B. Tone⁴; Bárbara Dritschel⁵; Alejandro Maiche¹; J. Douglas Steele⁶

¹ CIBPsi, Faculty of Psychology, Universidad de la República, Uruguay

² Medical Physics, NHS Tayside, University of Dundee, UK

³ Aberdeen Biomedical Imaging Centre, University of Aberdeen, U.K

⁴ Department of Psychology, Georgia State University, US

⁵ Department of Psychology, University of St. Andrews, U.K.

⁶ Division of Neuroscience, Medical Research Institute, University of Dundee, UK

Funding/Support: This study was funded by the University of Dundee, the Scottish Mental Health Research Network, Dr. Kathleen White's Clinical Neuroscience Research Endowment Fund and a postdoctoral fellowship from ANII (National Agency for Investigation and Innovation, Uruguay) to VBG. The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

1 Corresponding author:
2 Dr. Victoria Gradin
3 Center for Investigation of Basic Psychology (CIBPsi)
4 Faculty of Psychology
5 Universidad de la República
6 Montevideo 11200
7 Tel: (598) 24008555-285/286
8 Fax: (598) 24008640
9 Email: victoriagradin@gmail.com
10

Abstract

Background. Depression is a disabling disorder that significantly impacts on the interpersonal functioning of individuals. However, little is known about the neural substrates of such difficulties. In the last few years neuroeconomics, which combines imaging with multiplayer behavioural economic paradigms, has been used to study the neural substrates of normal and abnormal interpersonal interactions.

Methods. This study used functional Magnetic Resonance Imaging (fMRI) to investigate neural activity in unmedicated depressed participants (n=25) and matched healthy controls (n=25). During scanning, participants played a behavioural economic game, the Prisoner's Dilemma. In this game, the participant and a co-player independently choose either to cooperate or not cooperate with each other.

Results. Depressed participants reported higher levels of negative feelings (betrayal, guilt) during the game than did controls. Neural activation was compared between 'imbalanced' events (when one of the players cooperated and the other defected ('CD' and 'DC')) and 'draw' events (when both players either cooperated or defected ('CC' and 'DD')). Participants preferentially activated the anterior insula and the dorsolateral prefrontal cortex (DLPFC), a region implicated in cognitive control and regulation of emotions. Importantly, compared to controls depressed participants showed reduced activation in the left DLPFC, with the extent of signal reduction correlating with increased self-report feelings of guilt associated with DC outcomes.

Conclusions. Our findings suggest that depression is associated with reduced activation of the DLPFC during social events that involve unreciprocated cooperation. This abnormality may underlie anomalies in cognitive control and top down regulation of emotions during challenging social exchanges.

1 **Introduction**

2 Depression is a common and disabling disorder that can profoundly affect how an
3 individual interacts with others. People experiencing depression report difficulties
4 maintaining and enjoying relationships, less supportive social networks, less active
5 social lives, excessive reassurance seeking, poor intimate relationships and in general,
6 more maladaptive and less satisfactory social interactions (Billings *et al.*, 1983,
7 Fredman *et al.*, 1988, Hirschfeld *et al.*, 2000, Papakostas *et al.*, 2004, Segrin, 2000).

8 Despite the importance of interpersonal difficulties in psychiatric disorders such
9 as depression, the neurobiology of such impairments remains largely understudied,
10 partly due to difficulties in recreating and quantifying interpersonal exchanges (King-
11 Casas and Chiu, 2012). In the last few years, neuroeconomic approaches (Glimcher and
12 Rustichini, 2004) which combine interactive behavioural economic tasks with
13 neuroimaging, have been used to study interpersonal functioning and its neural
14 substrates in clinical populations (Gradin *et al.*, 2014, Hasler, 2011, King-Casas and
15 Chiu, 2012, King-Casas *et al.*, 2008, McClure-Tone *et al.*, 2011, Rilling *et al.*, 2007).
16 Behavioural economic tasks involve multiplayer interactive scenarios that allow
17 quantification of social exchanges and the study of social concepts such as fairness,
18 cooperation, and trust.

19 A well known behavioural economic task is the Prisoner's Dilemma (PD) game
20 (Axelrod and Hamilton, 1981). The PD allows examination of social relationships based
21 on cooperative and uncooperative behaviors. In each round two players independently
22 choose to either 'cooperate' with or 'defect' from each other. There are four possible
23 outcomes: both players cooperate (CC), one of the players cooperates and the other
24 defects (CD, DC) or both players defect (DD). Depending on the outcome, each player
25 is awarded a sum of money (see the payoff matrix, Fig. 1a).

Neuroimaging studies using economic behavioural tasks have reported that *positive* social exchanges activate regions of the reward circuitry, much as non-social rewards do. For example, studies using the Ultimatum Game—a task where participants accept or reject monetary offers made by others—have shown that fair offers typically activate the striatum in the receiver (Crockett *et al.*, 2013, Gradin *et al.*, 2014, Tabibnia *et al.*, 2008). Similarly, studies using the PD have reported striatal activation in response to reciprocated cooperation (Rilling *et al.*, 2002). In contrast, unfair offers during the Ultimatum Game have been reported to activate regions implicated in processing aversive emotions and salience detection (anterior insula), cognitive conflict (dorsal anterior cingulate and dorsomedial prefrontal cortex), cognitive control and emotion regulation (Dorsolateral Prefrontal Cortex (DLPFC))(Sanfey *et al.*, 2003). Unreciprocated cooperation during the PD has also been linked to anterior insula activation (Rilling *et al.*, 2008). In addition, it has been reported that the PD is associated with increased activation of the DLPFC when compared to the ‘stag hunt game’, possibly because the PD places higher cognitive control demands on participants (Emonds *et al.*, 2012), although other work has shown that the stag hunt game can involve relatively demanding mental computations (Yoshida *et al.*, 2008, Yoshida *et al.*, 2010).

Human studies of brain function in depression have reported a number of abnormalities. First, several studies indicate reduced activation in reward-related brain regions, particularly the striatum, in depression; this reduction may be linked to anhedonia (Eshel and Roiser, 2010, Gradin *et al.*, 2011, Zhang *et al.*, 2013). In addition, a recent study using the Ultimatum Game reported reduced striatal responses to increasing fairness of offers in depression, suggesting diminished responsiveness not only to material rewards but also to social rewards (Gradin *et al.*, 2014).

Depression has also been linked to emotion regulation models. These models hypothesize that depression is associated with hyperactivity of limbic regions that are involved in detecting emotions (bottom-up processes), and also with abnormal functioning of regions higher in the cognitive hierarchy, such as the DLPFC, resulting in abnormalities in control and regulation of emotions (top-down processes) (Disner *et al.*, 2011, Gotlib and Hamilton, 2008, Rive *et al.*, 2013).

In this study, we investigated neural activation in unmedicated depressed participants and healthy controls whilst they played the PD. Based on evidence that neural dysfunction in reward-related regions characterizes depressed individuals, we first hypothesized that depressed participants would show diminished striatal responses to reciprocated cooperation in comparison to controls. Second, based on evidence of neural anomalies in regions implicated in both bottom-up and top-down emotion processes in depressed individuals, we predicted that unreciprocated cooperation (CD) would be associated with enhanced activity in regions such as the insula, that are involved in processing emotionally salient aversive stimuli, and abnormal activity in regions such as the DLPFC, that are involved in top-down regulation of emotions.

Finally, we investigated emotional and neural responses during outcomes in which the participant defected while the co-player cooperated (DC). While this type of outcome may trigger some positive emotions (as the participant receives the highest payoff) it may also trigger the negative emotion of guilt (Rilling *et al.*, 2007). Since excessive feelings of guilt are a core symptom of depression (American Psychiatric Association, 2013), it was hypothesized that DC outcomes would lead to enhanced feelings of guilt in depression. At the neural level, it was hypothesized that, like unreciprocated cooperation (CD), DC outcomes could be associated with hyperactivity

- 1 of emotion detection regions and with abnormal activation of control and emotion
- 2 regulation regions such as the DLPFC.
- 3
- 4

Method

Participants

The study was approved by the local Research Ethics Committee and written informed consent was obtained from all participants. Data were acquired from 25 participants meeting criteria for an episode of DSM-IV depression and 25 healthy controls. The study was advertised within the Universities of Dundee and St. Andrews, UK. Potential participants were invited to self-nominate either for the depression or control group. Applicants were invited to a recruitment session (approximately 3-7 days before scanning) and were screened for depression and other psychiatric symptoms using the Mini International Neuropsychiatric Interview (MINI Plus V. 5.0) and symptom burden quantified using the Beck Depression Inventory (BDI, Beck *et al.*, 1961). Inclusion criteria for the depression group were: satisfying DSM-IV criteria for a major depressive disorder plus a score ≥ 16 in the BDI and at least 3 weeks of not taking antidepressant medication. Participants in the control group had no current or past history of depression or any other psychiatric disorder.

Participants in the depression and control groups were matched on the basis of gender, age, years of education, and estimated pre-morbid IQ according to the National Adult Reading Test (NART) (Nelson and Wilson, 1991) (Table 1).

Clinical ratings

Prior to scanning participants were assessed for symptom severity. Participants completed the (BDI, Beck *et al.*, 1961), the Hamilton Depression/Anxiety scale (HAM-D/A, Hamilton, 1959, Hamilton, 1960), the Montgomery-Asberg Depression Rating Scale (MADRS, Montgomery and Asberg, 1979), the Spielberger State Anxiety scale (Spielberger, 1983), the Rosenberg Self-Esteem Scale (RSES, Rosenberg, 1965), the

1 Positive Affect Negative Affect Scale (PANAS, Watson *et al.*, 1988) and the Snaith-
2 Hamilton hedonia scale (Snaith *et al.*, 1995). The HAM-D/A and the MADRS were
3 undertaken by a rater (VBG). Between the recruitment and scanning sessions,
4 participants completed the Sociotropy-Autonomy Scale (SAS, Beck *et al.*, 1983), the
5 Personal Style Inventory (PSI, Robins *et al.*, 1994), the Childhood Trauma
6 Questionnaire (CTQ, Bernstein *et al.*, 2003) and the Inventory of Interpersonal
7 Problems (IIP, Horowitz *et al.*, 1993).

9 ***Prisoner's Dilemma***

10 While in the scanner, participants played the PD game (Fig. 1a). Before scanning,
11 participants were shown how to play the PD (Supplementary Material). Participants
12 were told that they would be playing a game with a co-player who was outside the
13 scanner room. It was explained that on each trial, both players would have to make
14 simultaneous and independent decisions regarding whether to cooperate or not
15 cooperate with each other. Depending on their decisions they would both receive
16 earnings on each round. If they both cooperated they would both earn two pounds; if
17 one cooperated and the other did not they would earn zero and three respectively; if
18 neither cooperated they would both earn one pound. Participants were told that at the
19 end of the game they would be paid a percentage of the money they had accumulated
20 during the game and that the other player would also be similarly paid. In reality,
21 participants played the PD against a pre-programmed algorithm (McClure *et al.*, 2007,
22 Rilling *et al.*, 2002). This deception was necessary in order to minimize differences
23 among participants in the experience of the PD whilst ensuring ecological validity.

24 The PD algorithm (McClure *et al.*, 2007) generates each response based on the
25 participant's choices on the prior two rounds (Supplementary Material). A higher

1 frequency of volunteer cooperation in the prior two rounds elicited a higher probability
2 for a cooperative response, whilst a higher frequency of participant defection in the
3 previous two rounds elicited a higher probability of a defection response. Following
4 McClure et al (2007), the algorithm was designed so that the participant would also
5 experience periodic defection or “betrayal”. Specifically, the algorithm had a 50%
6 chance of defecting after four consecutive mutual cooperation trials. This effect was
7 introduced as previous PD studies have reported that participants otherwise engage in
8 mutual cooperation during much of the game (Rilling *et al.*, 2002). This pattern of play
9 would prevent participants from experiencing cooperation-defection outcomes in an
10 adequate number of trials for statistical analysis. Participants played two sessions of the
11 PD in the scanner. Each session lasted ~11.5 minutes and had 38 trials. The inter-trial
12 timing variation (‘jitter’) was determined using ‘Optseq’
13 (<http://surfer.nmr.mgh.harvard.edu/optseq/>).

14 After scanning, participants completed a questionnaire that assessed their
15 perceptions and emotional reaction to each of the PD outcomes (Supplementary
16 Material). Specifically, participants rated on nine-point Likert scales their satisfaction
17 with their earnings, as well as their feelings of cooperativeness, anger, betrayal and
18 guilt.

19 After the experiment, participants were debriefed regarding the cover story. All
20 participants believed the cover story. No participants reported being unhappy regarding
21 the deception. In the scanner, before playing the PD, participants played another
22 behavioural economic task, the Ultimatum Game (Gradin *et al.*, 2014). Participants
23 were paid according to their earnings in both games with an average of £17.

Behavioural and emotional analysis

Emotional ratings were analyzed using a three-way ANOVA with factors emotion, PD outcome and group. An ANOVA with factors outcome and group was used to analyze the number of occurrences of each outcome type as well as transition probabilities. The Greenhouse-Geisser correction was used for non-sphericity.

Neuroimaging analysis

For blood oxygen level dependent (BOLD) response imaging, T2* weighted gradient echo planar images were obtained using a 3T Siemens Magnetom Trio Tim MRI scanner with a 12-channel head coil (see the Supplementary Material for further details on data acquisition and preprocessing). SPM8 (<http://www.fil.ion.ucl.ac.uk/spm>) was used for analyses.

For the first level analysis, an event related design was used which modelled neural activation at the decision (when the participant selected a column of the payoffs matrix) and outcome (when the feedback screen with the final payoff was presented) times. Specifically, six regressors were defined: decision C, decision D, and outcomes CC, CD, DC and DD. Six head motion realignment parameter estimates were included as covariates of no interest. Regressors of interest were convolved with the SPM8 haemodynamic response function without time or dispersion derivatives. Contrast images of interest were taken to second level analyses and *within* and *between* group activations explored using one-sample and two-sample t-tests.

For the depression group, we tested for significant correlations between neural activity and self reported emotional rating scores. This correlational analysis was limited to the regions of interest where activation differed between groups. The

1 dependent variable in this analysis was the mean value of the parameter estimates across
2 voxels within the regions that showed between-group differences.

3 Unless otherwise stated, all analysis regions are reported as significant at a
4 whole brain $p < 0.05$ cluster level. This was achieved by using parameters identified
5 with Monte Carlo simulations: a simultaneous requirement for a voxel threshold of
6 $p < 0.01$ and a minimum cluster size of 68 continuous voxels (Slotnick *et al.*, 2003). All
7 images are presented at this threshold.

Results

Two control and three depression data sets were excluded from analyses as these participants did not experience all four PD outcome types during each scanning session.

Clinical ratings

Depressed participants scored higher than controls on measures of depression (BDI, HAM-D, MADRS), anxiety (HAM-A, Spielberger State Anxiety scale), negative affect (PANAS), and anhedonia (Snaith-Hamilton scale). The depression group also scored significantly higher in sociotropy (SAS, PSI), autonomy (PSI), child abuse and neglect (CTQ) and interpersonal problems (IIP). The depressed group had lower mean scores than controls on measures of self-esteem (RSES) and positive affect (PANAS). (see Table 1 for all comparisons).

Emotional responses

After scanning, participants rated their perceptions of and emotional reactions to each of the PD outcomes. A three-way ANOVA identified significant main effects for emotion ($F_{(2.18,93.77)} = 85.99, p < 0.001$), outcome ($F_{(2.5,107.50)} = 5.186, p = 0.004$), a significant emotion*group interaction ($F_{(2.18,93.77)} = 6.83, p = 0.001$) and a significant emotion*outcome interaction ($F_{(6.30,270.91)} = 65.18, p < 0.001$). Follow up analyses included independent ANOVAS for each emotion category with factors outcome and group. Each outcome type was associated with specific emotional reactions (Fig.1b, Table S1), consistent with previous work (Rilling *et al.*, 2007). Specifically, CC outcomes were associated with satisfaction with earnings and feelings of cooperativeness; CD outcomes with feelings of anger and betrayal; DC outcomes with guilt; DD outcomes with intermediate levels of all emotions.

For satisfaction with earnings there was a significant effect of group, with depressed participants reporting less satisfaction than controls ($F_{(1,43)}= 10.67, p=0.002$). Regarding feelings of cooperativeness and anger, there was no significant effect of group or a significant interaction with outcome type. For betrayal there was a significant effect of group ($F_{(1,43)}= 5.47, p=0.024$), with depressed participants reporting higher levels of betrayal than controls. There was also a significant group*outcome-type interaction ($F_{(2,3,99.52)}= 3.46, p=0.029$). Exploration of this interaction indicated that depressed participants reported significantly more betrayal than controls on DD outcomes ($p=0.004$); and there were no significant between-group differences on any other outcomes. Finally, for guilt there was a significant effect of group ($F_{(1,43)}= 5.54, p=0.023$), with depressed participants reporting higher levels than controls. There was also a non-significant interaction ($F_{(2.44,104.87)}= 2.45, p=0.08$), which might be considered a trend. Decomposition of this interaction indicated that depressed participants reported higher levels of guilt than controls on DC outcomes ($p=0.022$), not differentiating on all other outcomes. In summary, depressed participants reported less positive and more negative feelings in response to the PD game than controls.

Behavioural analyses

A mixed ANOVA with factors outcome and group was used to analyze the number of occurrences of each outcome type. There was a significant effect of outcome ($F_{(1.62,69.64)}= 18.27, p<0.001$), with CC and DD outcomes occurring more frequently than CD and DC outcomes. There were no significant group or interaction effects. We also analyzed transition probabilities (i.e. the probability of cooperating following a specific outcome in the previous trial). This analysis identified a significant effect of outcome ($F_{(2.58,110)}= 50.18, p<0.001$), with participants being more likely to cooperate

after CC outcomes, followed by DC, CD, and DD outcomes, respectively. There was no significant group or interaction effect (Supplementary Table S1). Controls and depressed participants did not differ on earnings during the game. We examined reaction times for cooperation and defection following co-player cooperation or defection and having group as a factor. This analysis yielded no significant main effect for group or significant interactions with the group factor.

Neuroimaging analyses

To detect brain regions involved in reward processing during the PD, we analyzed the contrast of reciprocated vs. unreciprocated cooperation (CC>CD) (Rilling *et al.*, 2004). For this contrast (Supplementary Table S2), across all participants (Fig. 2a) and also in the control group alone (Fig. 2b), we found activations extending through the nucleus accumbens and dorsal caudate, consistent with previous studies (Rilling *et al.*, 2002, Rilling *et al.*, 2004). At the same level of significance, no activation was observed in the striatum in the depression group (Fig. 2c), nor were there significant between-groups differences in this region. Similar results were obtained when considering the contrast [(CC+DC)>(CD+DD)] (i.e. every time a co-player cooperates vs. every time a co-player does not cooperate) in order to examine responses to rewarding feedback versus unrewarding feedback during the task (Supplementary Fig. S1).

We also examined brain activity in response to unreciprocated vs. reciprocated cooperation (CD>CC). Across all participants this analysis yielded activity in the bilateral DLPFC and the left anterior insula. No significant between-group differences in activation were observed in a priori regions of interest (Supplementary Fig. S2,

Supplementary Table S3). Next, we examined activation associated with events in which the co-player cooperated while the participant defected; mutually cooperative trials served as a baseline (contrast: DC>CC). Across all participants this contrast showed significant activation in the bilateral DLPFC and bilateral insula (Supplementary Fig. S3, Supplementary Table S4), with no significant between-group differences observed in any regions of interest.

As both CD and DC outcomes activated a network comprising the DLPFC and insula, we pooled these events in a single contrast [(CD+DC)>(CC+DD)] (Rilling *et al.*, 2002). That is, we compared the outcomes in which one of the players did not cooperate vs. the outcomes in which both players either cooperated or defected. As noted in prior research (Rilling *et al.*, 2002), CD and DC outcomes are typically aversive to at least one of the players and so are unlikely to be repeated, while CC and DD outcomes are more likely to repeat in a stable manner. Across all participants, this contrast elicited significant activations in the bilateral DLPFC, bilateral anterior insula and dorsomedial prefrontal cortex (Table 2, Fig. 3a). Controls also showed activations across these regions (Fig. 3b). Importantly, while depressed participants did show activation in the same network (Fig. 3c), they showed significantly diminished activation in the left DLPFC (Fig. 3d). This between-groups difference in the left DLPFC was driven by the combination of reduced responses to CD and DC outcomes in the depressed group (Fig. 3e). Given the evidence supporting a role for the DLPFC in regulating emotions (Rive *et al.*, 2013) we investigated whether activity in the left DLPFC correlated with self reported emotional ratings in response to CD (anger and betrayal) and DC (guilt) outcomes. This analysis showed that in the depression group, diminished activation in the left DLPFC correlated with increasing self-reported ratings of guilt in response to

1 DC outcomes ($r_{(22)}=-0.42, p=0.05$), Fig. 3f). No significant correlations were found for
2 anger and betrayal ratings.
3
4

Discussion

This study investigated behavioural, emotional, and neural responses during the PD game in adults with unmedicated depression.

At a neural level, it was found that during imbalanced (CD and DC) versus draw outcomes (CC and DD), depressed volunteers showed diminished activation in the left DLPFC compared to controls. The DLPFC has been implicated in processes of reasoning and higher cognition such as working memory, cognitive control (D'Esposito and Postle, 2015, Miller and Cohen, 2001), and also in the regulation of emotions (Okon-Singer *et al.*, 2015). Of relevance here, the DLPFC has been found to activate in response to unfair relative to fair offers during the Ultimatum Game (Sanfey *et al.*, 2003). This preferential DLPFC activation was interpreted as relating to the higher cognitive demands imposed by the unfair vs. fair offers (Sanfey *et al.*, 2003). Similarly, the DLPFC activation observed in our study was specifically associated with CD and DC outcomes. While CC and DD outcomes represent a draw, are stable, and tend to be repeated (Rilling *et al.*, 2002), CD and DC outcomes are associated with negative emotions (anger and betrayal in one case, guilt in the other) and are more likely to lead to altered behavior. If after a block of mutual cooperation the participant finds herself with a CD outcome, she may be more likely to choose to defect. Analogously, if after consecutive DD trials a DC outcome occurs, the participant will have to decide whether to follow the co-player signal and move to cooperation. Thus, the DLPFC activation found during these events may relate to higher cognitive demands placed by these outcomes in terms of emotion regulation and decision making.

Within this framework, diminished DLPFC activation during CD and DC outcomes in depressed volunteers suggests fewer cognitive resources for dealing with these events in terms of emotion regulation and decision making. Abnormal functioning

1 of the DLPFC in depression is consistent with previous findings. Depression has been
2 associated with reduced gray matter volume (Li *et al.*, 2010) and abnormally low levels
3 of resting state activity (Galynker *et al.*, 1998, Mayberg *et al.*, 1999) in the DLPFC. It
4 has also been reported that damage to the DLPFC confers vulnerability to depression
5 (Koenigs *et al.*, 2008).

6 It has been hypothesized, that abnormal functioning of the DLPFC in depression
7 may be associated with dysfunction in top-down regulation of emotion (Disner *et al.*,
8 2011, Gotlib and Hamilton, 2008). Research using several paradigms lends support to
9 this perspective. For example, one study showed decreased DLPFC activation in
10 depression while participants had to ignore fear stimuli, as well as on post-error trials,
11 suggesting impaired top-down control over affective interference and an impairment in
12 making post-error cognitive adjustments (Fales *et al.*, 2008). A second study found
13 decreased DLPFC activation during reversal learning in depression (Remijnse *et al.*,
14 2009). In a third study, participants with a history of depression failed to activate the
15 DLPFC when they heard critical remarks from their own mothers (Hooley *et al.*, 2005).

16 Findings do not point uniformly, however, to a consistent association between
17 depression and attenuated DLPFC activity; indeed, several studies have yielded
18 evidence of DLPFC hyperactivity in depression (Etkin and Schatzberg, 2011, Frodl *et*
19 *al.*, 2009, Strigo *et al.*, 2008). According to a recent review on emotion regulation (Rive
20 *et al.*, 2013), whether the DLPFC overactivates or underactivates in depression depends
21 on whether the emotion regulation process occurs in an automatic or voluntary manner.
22 Studies using tasks that engage *automatic* emotion regulation (Etkin and Schatzberg,
23 2011, Frodl *et al.*, 2009) have reported hyperactivity of the DLPFC in depression,
24 possibly related to the need for additional resources in order to override strong bottom-
25 up emotional influences. In contrast, studies using tasks that demand *voluntary* emotion

1 regulation (Fales *et al.*, 2008, Remijnse *et al.*, 2009), reported decreased DLPFC
2 activity in depression, suggesting a failure in recruitment of cognitive resources for
3 cognitive control and regulation.

4 Rive and colleagues (Rive *et al.*, 2013) have proposed that during early
5 automatic stages of emotion regulation, depressed subjects may be capable of regulating
6 emotions, but only with the recruitment of additional lateral prefrontal regions.

7 However, during explicit voluntary control, when the emotional experience is already
8 ongoing, this strategy of additional recruitment may fail, as reflected by abnormally
9 reduced activity in lateral prefrontal cortices. Studies of voluntary emotion regulation
10 have used tasks that involve learning from feedback or reappraisal (Rive *et al.*, 2013).

11 In our study, the PD implies learning from feedback which may be consistent with
12 reports of diminished DLPFC activation in depression feedback studies (Fales *et al.*,
13 2008, Remijnse *et al.*, 2009).

14 Cognitive theories of depression (Beck, 1979) propose that a core feature of the
15 illness is a bias towards negativity in the processing of information, with depressed
16 individuals selectively attending to and encoding negative events while filtering out
17 positive information. This bias may decrease the experience of positive emotions while
18 enhancing the feeling of negative emotions (Disner *et al.*, 2011). Consistent with this, in
19 our study depressed volunteers reported decreased satisfaction with earnings, as well as
20 increased feelings of betrayal and guilt in response to the PD game. Our finding of
21 heightened negative emotions in depressed participants is consistent with two previous
22 PD studies. In one study, it was found that depressed participants reported feelings of
23 self-devaluation, sadness and helplessness regarding exchanges during a modified
24 version of the PD (Hokanson *et al.*, 1980). In a second study (McClure *et al.*, 2007), it
25 was found that adolescent girls with anxiety and/or depression reported higher levels of

1 anger towards the co-player. It is possible that the reduced DLPFC activation observed
2 in the depression group underlies abnormalities in emotion regulation leading to the
3 observed enhanced negative feelings. Consistent with this, reduced left DLPFC in
4 depression correlated with increased feelings of guilt in response to DC outcomes.

5 It was not observed that depressed participants differed from controls in reward
6 related activation in the striatum in response to mutual cooperation. While depressed
7 volunteers had a weaker striatal response to mutual cooperation than controls, the
8 between-group difference in this region did not pass our significance threshold. Of note,
9 a previous study using the same participants as in the current study (Gradin *et al.*,
10 2014), showed significantly diminished striatal activation in response to increasing
11 fairness of offers during the Ultimatum Game in depression. Larger studies should
12 investigate reward-linked brain activation in depression using the PD and other
13 interactive paradigms. Similarly, depressed participants did not differ from controls in
14 emotion/salience detection regions such as the insula. As above, further work needs to
15 address the function of these regions in depression in the context of social interaction
16 paradigms.

17 Of note, while depressed participants differed from controls in emotional and
18 neural responses to the PD, the two groups did not differ in behavior. Two previous PD
19 studies have examined the behavior of depressed populations. One study (Hokanson *et*
20 *al.*, 1980) used a modified version of the PD in which each player's relative power was
21 manipulated. Results showed that when depressed individuals were in a controlling role,
22 the pattern of play in the PD was relatively exploitive and non-cooperative. In contrast,
23 another study using the PD (McClure *et al.*, 2007) found that adolescents with
24 anxiety/depression were more likely than controls to cooperate following co-player
25 cooperation, suggesting a stronger need for maintenance of positive social interactions.

1 Similarly to what is observed using the PD, studies using the Ultimatum Game have
2 shown inconsistent results in depression reporting either increased, decreased or
3 unchanged rejection rates to unfair offers (Destoop *et al.*, 2012, Gradin *et al.*, 2014,
4 Harle *et al.*, 2010, Scheele *et al.*, 2013). As has been noted (Gradin *et al.*, 2014, Pulcu
5 and Elliott, 2015, Wang *et al.*, 2015), these studies indicate that is not simple to predict
6 depressed behavior in the context of economic social exchange paradigms, and that
7 further work is needed in order to investigate whether specific depression subtypes can
8 be characterized by more consistent patterns of behavior.

9 A possible limitation of the study relates to the use of a university sample which
10 may limit generalizability of the results. This recruitment method was applied in order
11 to facilitate recruitment of unmedicated depressed participants, avoiding a potential
12 medication confound.

13 In summary, this study investigated patterns of emotional, behavioural, and
14 neural responses in unmedicated depressed and control participants during social
15 exchanges in the PD. In comparison to controls, the depressed group reported decreased
16 levels of satisfaction with earnings and increased levels of betrayal and guilt feelings.
17 Depressed participants also showed diminished DLPFC activation during exchanges in
18 which one player cooperated and the other defected versus the events in which both
19 players cooperated or defected. This abnormality in the DLPFC of depressed
20 individuals may contribute to impairments in cognitive control and top down regulation
21 of emotion during social situations that involve unreciprocated cooperation.

Acknowledgments

We would like to thank Christine Matthews, Mairi Stirling, Craig Adams and Fiona Grant for their support with obtaining ethics approval, the recruitment and data management. We also thank the staff from the Advanced Intervention Service (AIS) for valuable discussions regarding the study design. Finally, we thank all the participants in this study.

Funding/Support: This study was funded by the University of Dundee, the Scottish Mental Health Research Network, Dr. Kathleen White's Clinical Neuroscience Research Endowment Fund and a postdoctoral fellowship from ANII (National Agency for Investigation and Innovation, Uruguay) to VBG. The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

Conflicts of Interest and Financial Disclosures:

The authors report no conflicts of interest. JDS has received research funding via an honorarium associated with a lecture from Wyeth and an unrestricted educational grant from Schering-Plough.

References

- American Psychiatric Association** (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. Washington DC.
- Axelrod R, Hamilton WD** (1981). The evolution of cooperation. *Science* **211**, 1390-6.
- Beck AT** (1979). *Cognitive Theory of Depression*. J. Wiley and Sons: New York.
- Beck AT, Epstein N, Harrison RP, Emery G** (1983). Development of the Sociotropy-Autonomy Scale: A measure of personality factors in psychopathology. . *Unpublished manuscript, Piladelphia: University of Pennsylvania*.
- Beck AT, Ward HC, Mendelson M** (1961). An inventory for measuring depression. *Arch Gen Psychiatry* **4**, 561-571.
- Bernstein DP, Stein JA, Newcomb MD, Walker E, Pogge D, Ahluvalia T, Stokes J, Handelsman L, Medrano M, Desmond D, Zule W** (2003). Development and validation of a brief screening version of the Childhood Trauma Questionnaire. *Child Abuse and Neglect* **27**, 169-90.
- Billings AG, Cronkite RC, Moos RH** (1983). Social-environmental factors in unipolar depression: comparisons of depressed patients and nondepressed controls. *J Abnorm Psychol* **92**, 119-33.
- Crockett MJ, Apergis-Schoute A, Herrmann B, Lieberman MD, Muller U, Robbins TW, Clark L** (2013). Serotonin modulates striatal responses to fairness and retaliation in humans. *Journal of Neuroscience* **33**, 3505-13.
- D'Esposito M, Postle BR** (2015). The cognitive neuroscience of working memory. *Annual Review of Psychology* **66**, 115-42.
- Destoop M, Schrijvers D, De Grave C, Sabbe B, De Bruijn ER** (2012). Better to give than to take? Interactive social decision-making in severe major depressive disorder. *Journal of Affective Disorders* **137**, 98-105.
- Disner SG, Beevers CG, Haigh EA, Beck AT** (2011). Neural mechanisms of the cognitive model of depression. *Nat Rev Neurosci* **12**, 467-77.
- Emonds G, Declerck CH, Boone C, Vandervliet EJ, Parizel PM** (2012). The cognitive demands on cooperation in social dilemmas: an fMRI study. *Soc Neurosci* **7**, 494-509.
- Eshel N, Roiser JP** (2010). Reward and punishment processing in depression. *Biol Psychiatry* **68**, 118-24.

Etkin A, Schatzberg AF (2011). Common abnormalities and disorder-specific compensation during implicit regulation of emotional processing in generalized anxiety and major depressive disorders. *American Journal of Psychiatry* **168**, 968-78.

Fales CL, Barch DM, Rundle MM, Mintun MA, Snyder AZ, Cohen JD, Mathews J, Sheline YI (2008). Altered emotional interference processing in affective and cognitive-control brain circuitry in major depression. *Biological Psychiatry* **63**, 377-84.

Fredman L, Weissman MM, Leaf PJ, Bruce ML (1988). Social functioning in community residents with depression and other psychiatric disorders: results of the New Haven Epidemiologic Catchment Area Study. *J Affect Disord* **15**, 103-12.

Frodl T, Scheuerecker J, Albrecht J, Kleemann AM, Muller-Schunk S, Koutsouleris N, Moller HJ, Bruckmann H, Wiesmann M, Meisenzahl E (2009). Neuronal correlates of emotional processing in patients with major depression. *World J Biol Psychiatry* **10**, 202-8.

Galynker, II, Cai J, Oongseng F, Finestone H, Dutta E, Sersen D (1998). Hypofrontality and negative symptoms in major depressive disorder. *Journal of Nuclear Medicine* **39**, 608-12.

Glimcher PW, Rustichini A (2004). Neuroeconomics: the consilience of brain and decision. *Science* **306**, 447-52.

Gotlib IH, Hamilton JP (2008). Neuroimaging and Depression. *Current Directions In Psychological Science* **17**, 159-163.

Gradin VB, Kumar P, Waiter G, Ahearn T, Stickle C, Milders M, Reid I, Hall J, Steele JD (2011). Expected value and prediction error abnormalities in depression and schizophrenia. *Brain* **134**, 1751-64.

Gradin VB, Perez A, MacFarlane JA, Cavin I, Waiter G, Engelmann J, Dritschel B, Pomi A, Matthews K, Steele JD (2014). Abnormal brain responses to social fairness in depression: an fMRI study using the Ultimatum Game. *Psychological Medicine*, 1-11.

Hamilton M (1959). The assessment of anxiety states by rating. *British Journal of Medical Psychology* **32**, 50-5.

Hamilton M (1960). Rating scale for depression. *Journal of Neurology, Neurosurgery & Psychiatry* **23**, 56-62.

Harle KM, Allen JJ, Sanfey AG (2010). The impact of depression on social economic decision making. *J Abnorm Psychol* **119**, 440-6.

Hasler G (2011). Can the neuroeconomics revolution revolutionize psychiatry?
Neurosci Biobehav Rev **36**, 64-78.

Hirschfeld RM, Montgomery SA, Keller MB, Kasper S, Schatzberg AF, Moller HJ, Healy D, Baldwin D, Humble M, Versiani M, Montenegro R, Bourgeois M (2000). Social functioning in depression: a review. *J Clin Psychiatry* **61**, 268-75.

Hokanson JE, Sacco WP, Blumberg SR, Landrum GC (1980). Interpersonal behavior of depressive individuals in a mixed-motive game. *J Abnorm Psychol* **89**, 320-32.

Hooley JM, Gruber SA, Scott LA, Hiller JB, Yurgelun-Todd DA (2005). Activation in dorsolateral prefrontal cortex in response to maternal criticism and praise in recovered depressed and healthy control participants. *Biological Psychiatry* **57**, 809-12.

Horowitz LM, Rosenberg SE, Bartholomew K (1993). Interpersonal problems, attachment styles, and outcome in brief dynamic psychotherapy. *Journal of Consulting and Clinical Psychology* **61**, 549-60.

King-Casas B, Chiu PH (2012). Understanding interpersonal function in psychiatric illness through multiplayer economic games. *Biol Psychiatry* **72**, 119-25.

King-Casas B, Sharp C, Lomax-Bream L, Lohrenz T, Fonagy P, Montague PR (2008). The rupture and repair of cooperation in borderline personality disorder. *Science* **321**, 806-10.

Koenigs M, Huey ED, Calamia M, Rayment V, Tranel D, Grafman J (2008). Distinct regions of prefrontal cortex mediate resistance and vulnerability to depression. *Journal of Neuroscience* **28**, 12341-8.

Li CT, Lin CP, Chou KH, Chen IY, Hsieh JC, Wu CL, Lin WC, Su TP (2010). Structural and cognitive deficits in remitting and non-remitting recurrent depression: a voxel-based morphometric study. *Neuroimage* **50**, 347-56.

Mayberg HS, Liotti M, Brannan SK, McGinnis S, Mahurin RK, Jerabek PA, Silva JA, Tekell JL, Martin CC, Lancaster JL, Fox PT (1999). Reciprocal limbic-cortical function and negative mood: converging PET findings in depression and normal sadness. *Am J Psychiatry* **156**, 675-82.

McClure-Tone EB, Nawa NE, Nelson EE, Detloff AM, Fromm SJ, Pine DS, Ernst M (2011). Preliminary findings: neural responses to feedback regarding betrayal and cooperation in adolescent anxiety disorders. *Dev Neuropsychol* **36**, 453-72.

McClure EB, Parrish JM, Nelson EE, Easter J, Thorne JF, Rilling JK, Ernst M, Pine DS (2007). Responses to conflict and cooperation in adolescents with anxiety and mood disorders. *J Abnorm Child Psychol* **35**, 567-77.

Miller EK, Cohen JD (2001). An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* **24**, 167-202.

Montgomery SA, Asberg M (1979). A new depression scale designed to be sensitive to change. *British Journal of Psychiatry* **134**, 382-389.

Nelson HE, Wilson JR (1991). *The revised national adult reading test - test manual*. NFER-Wilson: Winsor.

Okon-Singer H, Hendler T, Pessoa L, Shackman AJ (2015). The neurobiology of emotion-cognition interactions: fundamental questions and strategies for future research. *Front Hum Neurosci* **9**, 58.

Papakostas GI, Petersen T, Mahal Y, Mischoulon D, Nierenberg AA, Fava M (2004). Quality of life assessments in major depressive disorder: a review of the literature. *Gen Hosp Psychiatry* **26**, 13-7.

Pulcu E, Elliott R (2015). Neural origins of psychosocial functioning impairments in major depression. *Lancet Psychiatry* **2**, 835-43.

Remijnse PL, Nielen MM, van Balkom AJ, Hendriks GJ, Hoogendijk WJ, Uylings HB, Veltman DJ (2009). Differential frontal-striatal and paralimbic activity during reversal learning in major depressive disorder and obsessive-compulsive disorder. *Psychological Medicine* **39**, 1503-18.

Rilling J, Gutman D, Zeh T, Pagnoni G, Berns G, Kilts C (2002). A neural basis for social cooperation. *Neuron* **35**, 395-405.

Rilling JK, Glenn AL, Jairam MR, Pagnoni G, Goldsmith DR, Elfenbein HA, Lilienfeld SO (2007). Neural correlates of social cooperation and non-cooperation as a function of psychopathy. *Biological Psychiatry* **61**, 1260-71.

Rilling JK, Goldsmith DR, Glenn AL, Jairam MR, Elfenbein HA, Dagenais JE, Murdock CD, Pagnoni G (2008). The neural correlates of the affective response to unreciprocated cooperation. *Neuropsychologia* **46**, 1256-66.

Rilling JK, Sanfey AG, Aronson JA, Nystrom LE, Cohen JD (2004). Opposing BOLD responses to reciprocated and unreciprocated altruism in putative reward pathways. *Neuroreport* **15**, 2539-43.

Rive MM, van Rooijen G, Veltman DJ, Phillips ML, Schene AH, Ruhe HG (2013). Neural correlates of dysfunctional emotion regulation in major depressive disorder. A

systematic review of neuroimaging studies. *Neuroscience and Biobehavioral Reviews* **37**, 2529-53.

Robins CJ, Ladd J, Welkowitz J, Blaney PH, Diaz R, Kutcher G (1994). The Personal Style Inventory: Preliminary Validation Studies of New Measures of Sociotropy and Autonomy. *Journal of Psychopathology and Behavioral Assessment* **16**.

Rosenberg M (1965). *Society and the adolescent self-image*. Princeton NJ: Princeton University Press.

Sanfey AG, Rilling JK, Aronson JA, Nystrom LE, Cohen JD (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science* **300**, 1755-8.

Scheele D, Mihov Y, Schwederski O, Maier W, Hurlemann R (2013). A negative emotional and economic judgment bias in major depression. *Eur Arch Psychiatry Clin Neurosci*.

Segrin C (2000). Social skills deficits associated with depression. *Clinical Psychology Review* **20**, 379-403.

Slotnick SD, Moo LR, Segal JB, Hart J, Jr. (2003). Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. *Brain Res Cogn Brain Res* **17**, 75-82.

Snaith RP, Hamilton M, Morley S, Humayan A, Hargreaves D, Trigwell P (1995). A scale for the assessment of hedonic tone the Snaith-Hamilton Pleasure Scale. *Br J Psychiatry* **167**, 99-103.

Spielberger CD (1983). *State-trait anxiety inventory for adults, STAI Form Y-1*. Mind Garden: Palo-Alto.

Strigo IA, Simmons AN, Matthews SC, Craig AD, Paulus MP (2008). Association of major depressive disorder with altered functional brain response during anticipation and processing of heat pain. *Archives of General Psychiatry* **65**, 1275-84.

Tabibnia G, Satpute AB, Lieberman MD (2008). The sunny side of fairness: preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychol Sci* **19**, 339-47.

Wang Y, Yang LQ, Li S, Zhou Y (2015). Game Theory Paradigm: A New Tool for Investigating Social Dysfunction in Major Depressive Disorders. *Front Psychiatry* **6**, 128.

Watson D, Clark LA, Tellegen A (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology* **54**, 1063-70.

- 1 **Yoshida W, Dolan RJ, Friston KJ** (2008). Game theory of mind. *PLoS Comput Biol* **4**,
2 e1000254.
- 3 **Yoshida W, Seymour B, Friston KJ, Dolan RJ** (2010). Neural mechanisms of belief
4 inference during cooperative games. *J Neurosci* **30**, 10744-51.
- 5 **Zhang WN, Chang SH, Guo LY, Zhang KL, Wang J** (2013). The neural correlates of
6 reward-related processing in major depressive disorder: a meta-analysis of functional
7 magnetic resonance imaging studies. *J Affect Disord* **151**, 531-9.

Table 1 Participant details

	Control	Depression	Significance
n	25	25	
Fem/Male	17/8	17/8	NS
Age	25.44±5.02	25.48±5.52	p=0.98, NS
NART	123.76±2.82	124.28±2.05	p=0.46, NS
Years of education	16.52±3.02	17.26±2.93	p=0.38, NS
BDI	0.40±0.76	28.80±9.06	p<0.001
HAM-D	0.16±0.47	12.44±4.23	p<0.001
MADRS	0.48±0.82	20.80±6.97	p<0.001
HAM-A	0.44±0.71	9.28±4.17	p<0.001
Spielberger State Anxiety	25.60±3.79	48.48±10.62	p<0.001
RSES	25.40±3.48	9.20±3.82	p<0.001
PANAS positive affect	38.96±4.29	18.24±4.78	p<0.001
PANAS negative affect	11.92±2.40	25.64±6.43	p<0.001
Snaith-Hamilton	4.12±3.40	20.12±4.53	p<0.001
SAS sociotropy	57.92±11.79	80.56±17.76	p<0.001
SAS autonomy	66.88±13.49	67.16±14.54	p=0.9, NS
PSI sociotropy	82.32±15.21	104.84±15.36	p<0.001
PSI autonomy	73.56±16.12	94.56±10.95	p<0.001
CTQ	5.68±0.96	9.67±2.75	p<0.001
IIP	54.84±28.16	110.36±27.31	p<0.001

Values are mean ± SD; NART, National Adult Reading Test; BDI, Beck Depression Inventory; HAM-D/A, Hamilton Depression/Anxiety scale; MADRS, Montgomery-Asberg Depression Rating Scale; RSES, Rosenberg Self-Esteem Scale; PANAS PA/NA, Positive Affect Negative Affect Scale; SAS, Sociotropy-Autonomy Scale; PSI, Personal Style Inventory; CTQ, Childhood Trauma Questionnaire; IIP, Inventory of Interpersonal Problems; *p*-values of the independent samples t-test are provided; NS, no significant difference between groups.

Table 2 Within group and between group brain activations during the outcomes in which one player cooperated while the other did not vs. the times in which both cooperated or defected (contrast [(CD+DC)>(CC+DD)])

	BA	Cluster size	x	y	z	T
<i>Activation for the contrast [(CD+DC)>(CC+DD)]</i>						
All subjects						
L frontal lobe, middle frontal gyrus	9	2615	-40	10	34	4.73
L anterior insula		''	-34	24	-4	6.65
R frontal lobe, precentral gyrus	9	2813	40	6	32	5.45
R anterior insula		''	32	24	-2	5.12
Frontal lobe, medial frontal gyrus	8	1589	-8	28	50	3.97
Superior midbrain, thalamus		2524	-4	-18	-4	5.22
L occipital lobe, superior occipital gyrus	19	1026	-40	-80	22	3.79
R occipital lobe, superior occipital gyrus	19	3045	34	-76	24	3.85
Control group						
L frontal lobe, precentral gyrus	9	2672	-40	14	40	6.41
L anterior insula		''	-36	24	-6	5.09
R frontal lobe, middle frontal gyrus	9	1708	42	8	44	5.14
R anterior insula		''	34	20	-6	4.69
Frontal lobe, medial frontal gyrus	9	844	2	48	34	4.02
Thalamus		2782	6	-24	2	7.24
Parietal lobe, precuneus	7	142	0	-70	46	3.69
Parietal lobe, precuneus	31	1623	-12	-72	20	3.62
R parietal lobe, superior parietal lobule	7	''	30	-64	56	3.29
Depression group						
L frontal lobe, inferior frontal gyrus	9	111	-42	4	32	3.63
R frontal lobe, inferior frontal gyrus	45,	357	50	24	24	3.79
L anterior insula		526	-38	18	-12	5.09
R anterior insula		110	32	28	0	3.24
Superior midbrain		228	6	-14	-8	4.94
R temporal lobe, fusiform gyrus	37	221	46	-54	-12	3.60
Control > Depression						
L frontal lobe, middle frontal gyrus	9	136	-36	24	32	3.20
Posterior thalamus		322	0	-28	6	4.04
Cerebellum		256	14	-54	-6	3.06
Depression > Control						
No significant activations						

Coordinates (x, y, z) reported in MNI space; R/L=right/left; BA=Brodmann area; '' indicates that the peak belongs to the same cluster as the peak above. All results significant at $p < 0.05$ cluster extent corrected across the whole-brain.

Figure legends

Fig. 1 The Prisoner's Dilemma (PD) game and emotional results

(a) On each trial the participant and a (supposed) co-player make a simultaneous and independent decision regarding whether to cooperate or not cooperate (defect) with each other. Depending on their decisions they receive a payoff. At the beginning of the trial the participant sees the payoff matrix displayed on the screen. The columns of the matrix represent the participant's choices and the rows correspond to the co-player's choices. Whether the cooperative or not cooperative choice appears in the left or right column was randomized across trials. In the payoff matrix, numbers in BOLD/light-gray correspond to the participant/co-player payoffs. Once the participant makes his/her choice the selected column of the matrix turns yellow. At the end of the trial, the payoff matrix is shown with only one cell highlighted, indicating the outcome of the trial. rt, reaction time; s, seconds (b) Emotional responses to each of the PD game outcomes. Error bars denote standard deviations

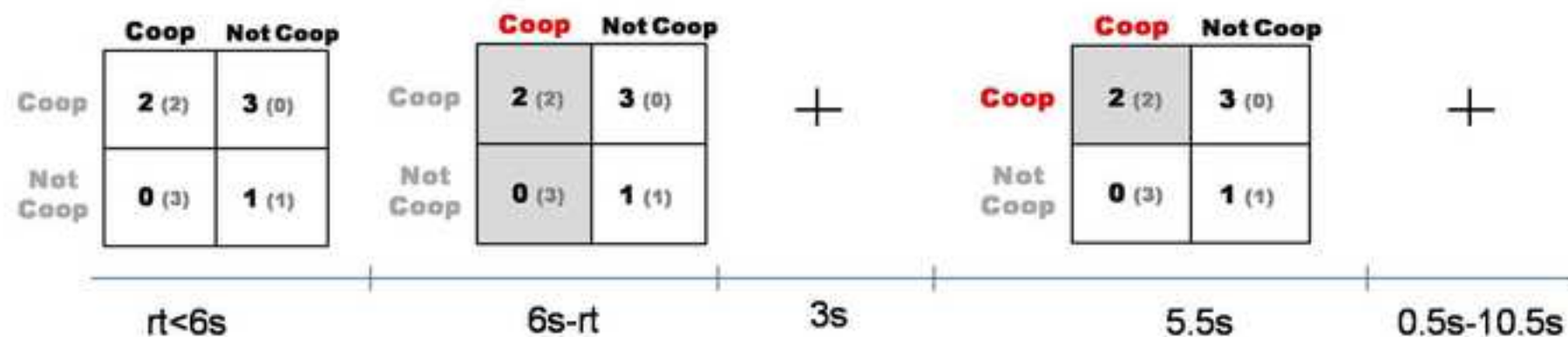
Fig.2 Neural responses to reciprocated vs. unreciprocated cooperation

Neural responses to reciprocated vs. unreciprocated cooperation (CC>CD) across all participants (a), in controls (b) and in depression (c).

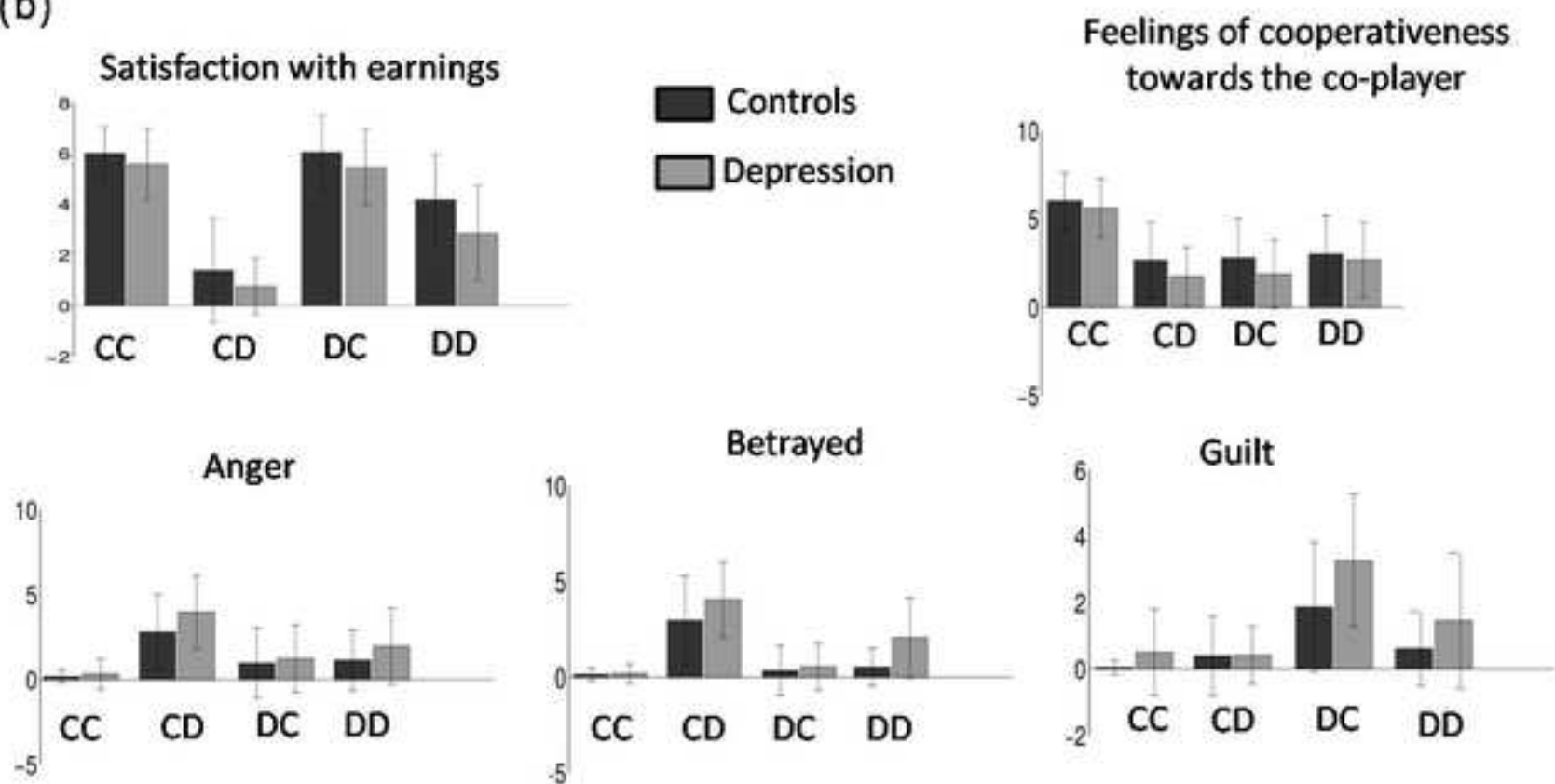
Fig.3 Neural responses during events in which one player cooperated while the other did not vs. events where both players cooperated or defected. (contrast: [(CD+DC)>(CC+DD)])

Neural responses across all participants (a), in controls (b) and in depression (c). Controls exhibited stronger responses in the left DLPFC than depressed participants (d). (e) Mean value of parameter estimates across voxels within a sphere of diameter 10mm centred at peak coordinates (-36 24 32) of the left DLPFC. Error bars denote standard error of the mean. (f) Correlation within the depression group. X axis: self reported feelings of guilt in response to DC outcomes during the PD game; Y axis: mean value of parameter estimates for the contrast [(CD+DC)>(CC+DD)] across voxels in the left DLPFC region where depressed participants differed from controls. DLPFC, dorsolateral prefrontal cortex.

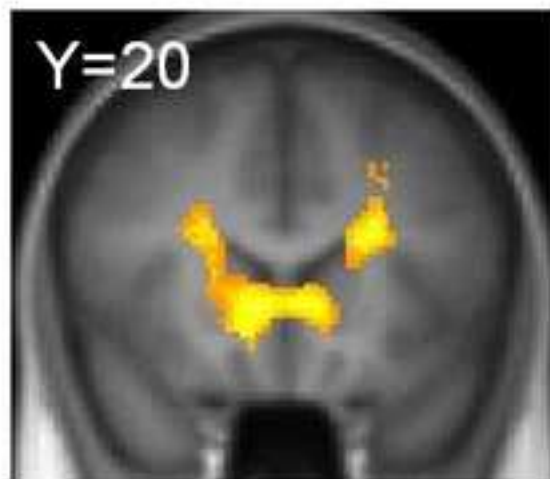
(a)



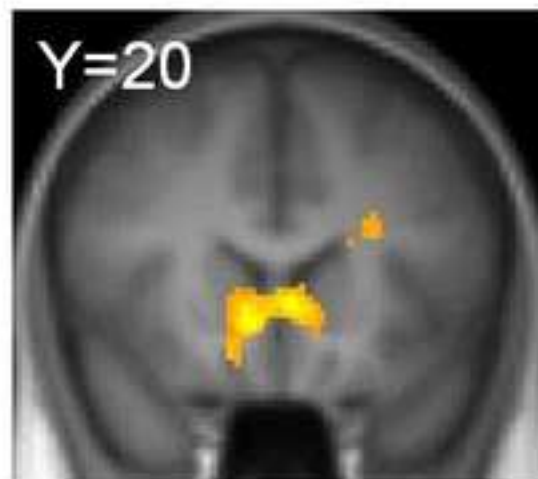
(b)



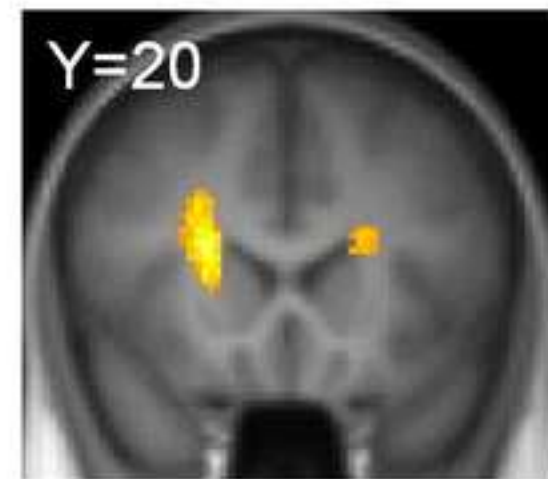
(a) All

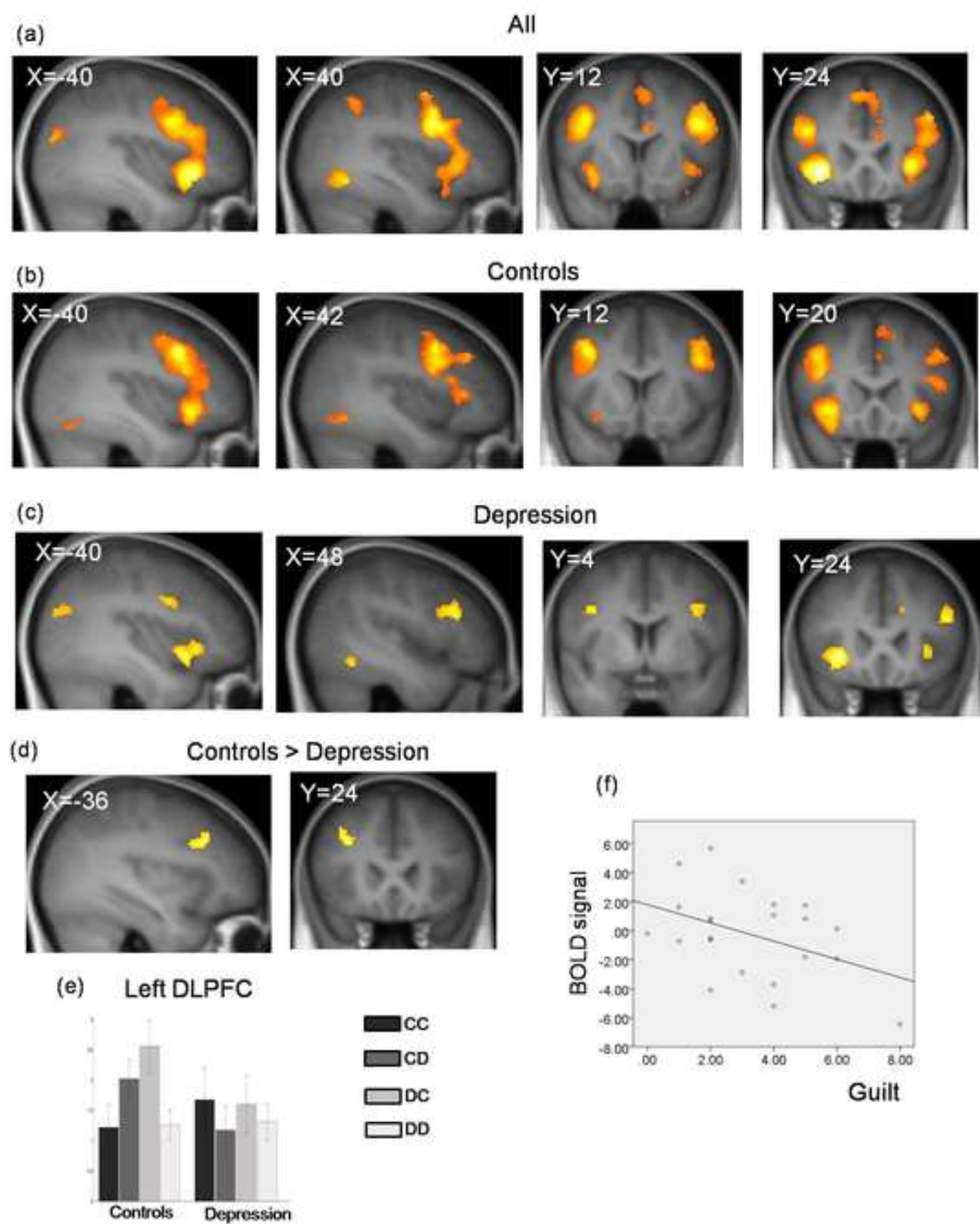


(b) Controls



(c) Depression





Online Supplementary Material

Instructions for the participant on how to play the Prisoner's Dilemma

These Instructions were explained at the same time as the participant was shown a figure with a display of the Prisoner's Dilemma screens.

“During this game you will be playing with a co-player who will be outside the scanner room. On each trial, both of you will have to decide whether to cooperate or not cooperate with each other. Both of you will be making your decisions at the same time. Your options will be represented as columns in this matrix while the options of your co-player will be represented as rows. The ‘cooperate’ and ‘not cooperate’ columns can be either at the left or right. If you want to choose the column at the left/right push the button at your left/right hand. Once you make an option, the chosen column will appear highlighted in yellow. Right after you will see the option made by your co-player and only the cell that represents the options of you both will remain highlighted. This cell shows the earnings that both of you will make on this trial. Your earnings are shown in bold numbers while your co-player earnings are shown in light grey. If you both cooperate you both earn 2; if you cooperate and your co-player does not cooperate you earn 0 and your co-player earns 3; if you do not cooperate and your co-player does cooperate you earn 3 and your co-player earns 0; if neither of you cooperate you will both earn 1. At the end of the game you will be paid a percentage of your total earnings. Your co-player will also be paid a percentage of his/her total earnings. The game will last for two sessions of 38 trials”.

After being explained the Instructions, the participant had a chance to practice a few trials of the game in a computer.

The Prisoner's Dilemma algorithm

During the Prisoner's Dilemma game, participants played against an algorithm implemented in Matlab that generates each response based on the outcomes of the two previous rounds (McClure *et al.*, 2007). So given a trial i , the algorithm would take into account [participant's choice (i-2), algorithm's choice (i-2), participant's choice (i-1), algorithm's choice (i-1)]

Specifically, the probability that the algorithm would cooperate on each trial was as follows:

- 1) Round 1: 100%
- 2) Round 2:
 - a. If Round 1 outcome was CC, then 93%
 - b. If Round 1 outcome was DC, then 36%
- 3) Rounds 3-19:
 - a. If outcome of prior 2 rounds was CCCC: 92%
 - b. If outcome of prior 2 rounds was CDCC: 86%
 - c. If outcome of prior 2 rounds was DCCC: 78%
 - d. If outcome of prior 2 rounds was DDCC: 50%
 - e. If outcome of prior 2 rounds was CCCD: 58%
 - f. If outcome of prior 2 rounds was CDCD: 0%
 - g. If outcome of prior 2 rounds was DCCD: 33%
 - h. If outcome of prior 2 rounds was DDCCD: 33%
 - i. If outcome of prior 2 rounds was CCDC: 86%
 - j. If outcome of prior 2 rounds was CDDC: 80%
 - k. If outcome of prior 2 rounds was DCDC: 33%
 - l. If outcome of prior 2 rounds was DDDC: 20%
 - m. If outcome of prior 2 rounds was CCDD: 50%
 - n. If outcome of prior 2 rounds was CDDD: 38%
 - o. If outcome of prior 2 rounds was DCDD: 50%
 - p. If outcome of prior 2 rounds was DDDD: 43%
- 4) Rounds 20-38:
 - a. If outcome of prior 2 rounds was CCCC: 92%
 - b. If outcome of prior 2 rounds was CDCC: 90%
 - c. If outcome of prior 2 rounds was DCCC: 100%
 - d. If outcome of prior 2 rounds was DDCC: 60%
 - e. If outcome of prior 2 rounds was CCCD: 13%
 - f. If outcome of prior 2 rounds was CDCD: 20%
 - g. If outcome of prior 2 rounds was DCCD: 67%
 - h. If outcome of prior 2 rounds was DDCCD: 33%
 - i. If outcome of prior 2 rounds was CCDC: 83%
 - j. If outcome of prior 2 rounds was CDDC: 63%
 - k. If outcome of prior 2 rounds was DCDC: 0%

- l. If outcome of prior 2 rounds was DDDC: 33%
- m. If outcome of prior 2 rounds was CCDD: 33%
- n. If outcome of prior 2 rounds was CDDD: 8%
- o. If outcome of prior 2 rounds was DCDD: 50%
- p. If outcome of prior 2 rounds was DDDD: 25%

The algorithm was also designed so that the participant would experience periodic defection. This was achieved by setting the condition that after four consecutive trials of mutual defection the algorithm had a 50% chance of defecting.

Questionnaire on emotional responses

After completing the scanning session, participants were invited to complete a questionnaire that assessed their perceptions and emotional reaction to each of the PD outcomes. Specifically, for each of the PD outcomes (CC, CD, DC, DD) participants rated on nine-point Likert scales their answers to the following questions:

- How satisfied were you with your earnings?
- How were your feelings of cooperativeness towards the co-player?
- How were your feelings of anger towards the other player?
- How betrayed did you feel?
- How guilty did you feel?

Supplementary methods on data acquisition and pre-processing

For blood oxygen level dependent (BOLD) response imaging, T2* weighted gradient echo planar images were obtained using a 3T Siemens Magnetom Trio Tim MRI scanner with a 12-channel head coil. A total of 37 sequential slices of 3.5 mm thickness and 0.5 mm slice gap were obtained for each volume. In order to minimize the susceptibility artefact, slice orientation was initially orientated parallel to the AC-PC line, then rotated 30 degrees towards the coronal plane for scanning. Two hundred and seventy six volumes were obtained with a TR of 2.5 s, TE 30 ms, flip 90°, FOV 224 mm and matrix 64x64. The first four volumes were discarded to allow for scanner transient effects.

SPM8 (<http://www.fil.ion.ucl.ac.uk/spm>) was used for analyses. The first image from each session was aligned to the first scan of the first session. Then the images from each session were aligned to the first image of the session. For each subject, the structural T1 image was coregistered to the average realigned image. The coregistered T1 image was used to derive parameters for spatial normalization to the SPM8 Montreal Neurological Institute (MNI) T1 template with the parameters applied to each fMRI time-series image. The resultant realigned and spatially normalized images were smoothed with an 8 mm FWHM Gaussian kernel.

Table S1 Emotional and behavioural responses to the Prisoner's Dilemma

	Control (mean \pm SD)	Depression (mean \pm SD)
Satisfaction with earnings		
CC	6.00 \pm 1.04	5.59 \pm 1.40
CD	1.39 \pm 2.06	0.77 \pm 1.11
DC	6.04 \pm 1.46	5.45 \pm 1.50
DD	4.17 \pm 1.77	2.86 \pm 1.88
Cooperativeness with co-player		
CC	6.00 \pm 1.60	5.59 \pm 1.65
CD	2.65 \pm 2.17	1.73 \pm 1.67
DC	2.78 \pm 2.24	1.86 \pm 1.93
DD	3.00 \pm 2.17	2.68 \pm 2.12
Anger at co-player		
CC	0.17 \pm 0.39	0.32 \pm 0.89
CD	2.78 \pm 2.21	3.95 \pm 2.17
DC	0.96 \pm 2.06	1.23 \pm 1.97
DD	1.13 \pm 1.79	1.95 \pm 2.24
Feelings of betrayal		
CC	0.13 \pm 0.34	0.18 \pm 0.50
CD	2.96 \pm 2.33	4.05 \pm 1.96
DC	0.35 \pm 1.30	0.55 \pm 1.22
DD	0.52 \pm 0.99	2.05 \pm 2.10
Feelings of guilt		
CC	0.04 \pm 0.21	0.50 \pm 1.30
CD	0.39 \pm 1.20	0.41 \pm 0.85
DC	1.87 \pm 1.96	3.27 \pm 2.00
DD	0.61 \pm 1.12	1.45 \pm 2.04
Average number of outcome types		
CC	20.48 \pm 8.46	20.27 \pm 9.65
CD	15.48 \pm 3.93	15.64 \pm 6.05
DC	14.70 \pm 5.70	13.32 \pm 5.02
DD	25.00 \pm 7.54	26.50 \pm 9.24
Transition probabilities (probability of cooperation after each outcome type)		
CC	0.72 \pm 0.28	0.73 \pm 0.27
CD	0.35 \pm 0.19	0.38 \pm 0.16
DC	0.49 \pm 0.26	0.50 \pm 0.23
DD	0.25 \pm 0.15	0.22 \pm 0.15

See the main text for details on the statistical analysis of these variables

Table S2 Within group brain activations for reciprocated vs. unreciprocated cooperation

	BA	Cluster size	x	y	z	T
<i>Activation to reciprocated vs. unreciprocated cooperation (CC>CD)</i>						
All subjects						
L nucleus accumbens and dorsal caudate		3659	-12	22	0	4.17
R nucleus accumbens and dorsal caudate		''	10	26	0	4.39
R frontal lobe, superior frontal gyrus	10	500	14	58	-2	3.93
L temporal lobe, superior temporal gyrus and posterior insula	22	827	-40	-2	14	3.82
R temporal lobe, superior temporal gyrus	22	114	60	-6	0	3.64
L parietal lobe, inferior parietal lobule	39	109	-50	-68	40	3.75
L frontal lobe, paracentral lobule	5	2703	-2	-32	54	3.11
Control group						
L nucleus accumbens and dorsal caudate		1039	-10	24	-2	4.14
R nucleus accumbens and dorsal caudate		''	2	20	0	3.78
Depression group						
R frontal lobe, medial frontal gyrus	10	244	14	60	4	3.45
L temporal lobe, middle temporal gyrus	21	2362	-60	-6	-4	4.26
L posterior insula		''	-38	-4	16	3.90
R posterior insula		524	44	-16	18	3.43

Coordinates (x, y, z) reported in MNI space; R/L=right/left; BA=Brodmann area; ''

indicates that the peak belongs to the same cluster as the peak above. All results

significant at $p < 0.05$ cluster extent corrected across the whole-brain.

Table S3 Within and between group brain activations for unreciprocated vs. reciprocated cooperation

	BA	Cluster size	x	y	z	T
<i>Activation to unreciprocated vs. reciprocated cooperation (CD>CC)</i>						
All subjects						
L frontal lobe, middle frontal gyrus	9	104	-42	10	32	2.96
R frontal lobe, inferior frontal gyrus	9	120	40	6	34	3.46
L anterior insula		381	-34	24	-4	4.29
L parietal lobe, superior parietal lobule	7	382	-30	-66	52	3.38
R parietal lobe, superior parietal lobule	7	1091	32	-82	22	4.79
Occipital lobe, cuneus	17	936	6	-72	8	4.31
Superior midbrain		394	-6	-16	-6	3.87
Cerebellum		92	-2	-40	-14	3.06
Control group						
R frontal lobe, inferior frontal gyrus	9	83	42	8	34	3.10
L anterior insula		106	-30	18	-16	3.29
Occipital lobe, cuneus	17	2316	8	-70	4	5.14
R occipital lobe, superior occipital gyrus	19	265	32	-82	22	4.11
L parietal lobe, superior parietal lobule	7	88	-30	-56	56	3.24
R parietal lobe, superior parietal lobule	7	197	28	-62	60	3.80
Superior midbrain		642	-4	-26	0	4.63
Depression group						
L anterior insula		189	-38	26	0	3.79
R parietal lobe, superior parietal lobule	7	136	12	-70	54	3.25
Control > Depression						
Posterior thalamus		107	0	-24	4	3.37
Cerebellum		116	4	-46	-28	4.02
Cerebellum		537	12	-62	-8	3.90
Depression > Control						
No significant activations						

Coordinates (x, y, z) reported in MNI space; R/L=right/left; BA=Brodmann area. All results significant at $p < 0.05$ cluster extent corrected across the whole-brain. Note that between group activations for the contrast (CC>CD) are the same as reported in this table for the contrast (CD>CC) but interchanging the labels “Control>Depression” and “Depression>Control”

Table S4 Within and between group brain activations during outcomes where the participant defected while the co-player cooperated vs. reciprocated cooperation (contrast DC>CC)

	BA	Cluster size	x	y	z	T
<i>Activation for the contrast (DC>CC)</i>						
All subjects						
L frontal lobe, middle frontal gyrus	9	1133	-44	16	32	3.59
R frontal lobe, precentral gyrus	9	1884	40	8	34	4.36
R anterior insula		''	34	24	0	4.57
L anterior insula		675	-32	20	-6	5.34
L parietal lobe, precuneus	39	212	-36	-70	32	3.04
R parietal lobe, precuneus	7	783	18	-74	50	3.75
Superior midbrain		312	-6	-16	-8	3.15
Cerebellum		''	2	-40	-14	3.53
Control group						
L frontal lobe, precentral gyrus	9	627	-40	14	38	3.52
R frontal lobe, inferior frontal gyrus	9	984	40	6	24	3.89
L anterior insula		291	-32	20	-6	3.91
R anterior insula		71	32	22	-2	3.27
Posterior thalamus		889	0	-22	4	5.85
R parietal lobe, inferior parietal lobule	40	260	34	-52	34	4.99
Depression group						
L anterior insula		486	-32	26	-2	5.15
R anterior insula		204	34	26	0	3.49
R temporal lobe, fusiform gyrus	37	161	40	-48	-16	4.89
R parietal lobe, superior parietal lobule	7	91	24	-66	46	3.47
Control > Depression						
Posterior thalamus and cerebellum		2433	-2	-26	4	4.98
Depression > Control						
No significant activations						

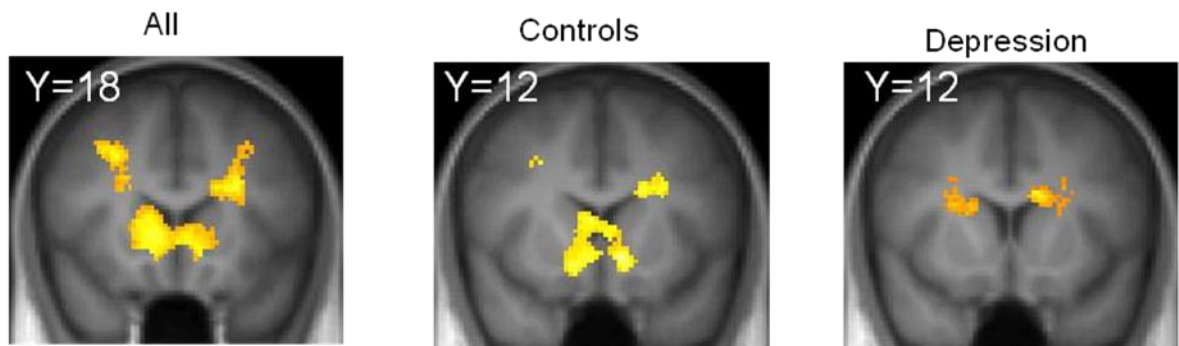
Coordinates (x, y, z) reported in MNI space; R/L=right/left; BA=Brodmann area; ''

indicates that the peak belongs to the same cluster as the peak above. All results

significant at $p < 0.05$ cluster extent corrected across the whole-brain.

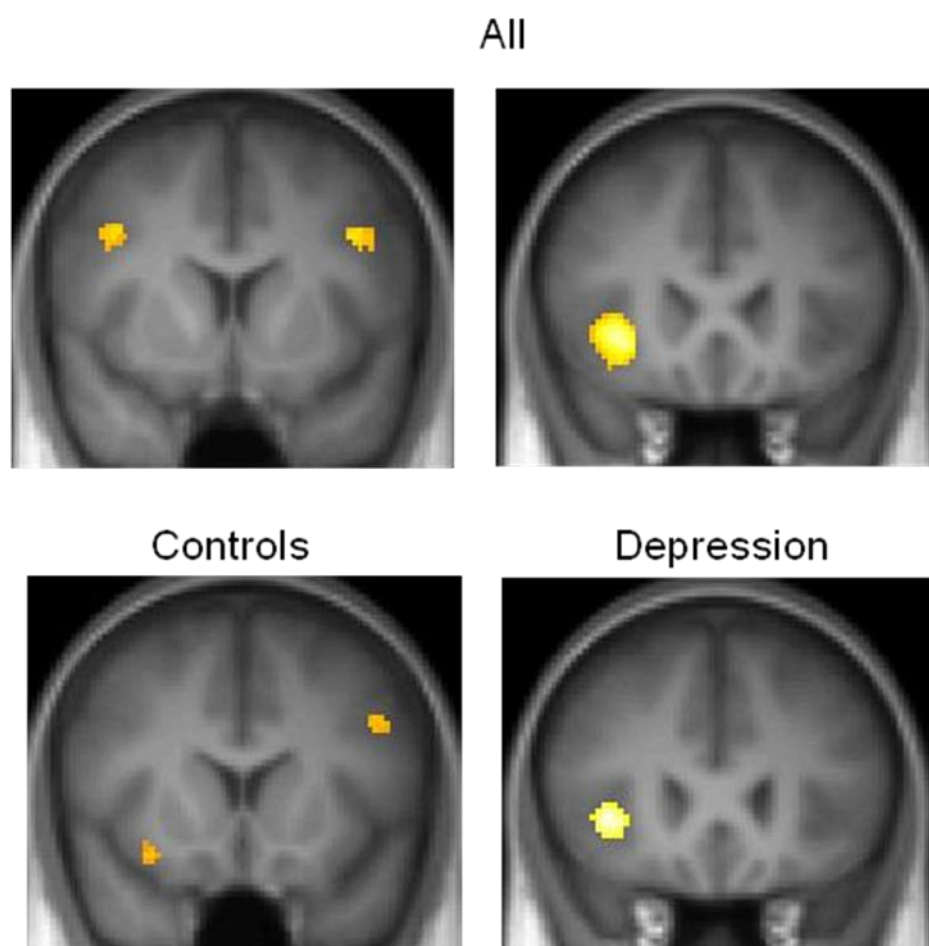
Supplementary figure S1

Neural responses for the contrast $[(CC+DC)>(CD+DD)]$ (i.e. every time that the co-player cooperates vs. every time that the co-player does not cooperate). For this contrast, across all participants $((6\ 22\ 0), t=4.09)$; $((-12\ 20\ 2), t=3.86)$ and in controls $((-8\ 2\ -4), t=4.32)$; $((8\ 6\ -6), t=4.06)$, activation was observed in the striatum. In the depressed group activation was not observed at our significance threshold. However, between groups differences in the striatum failed to reach significance, again at our chosen significance threshold.



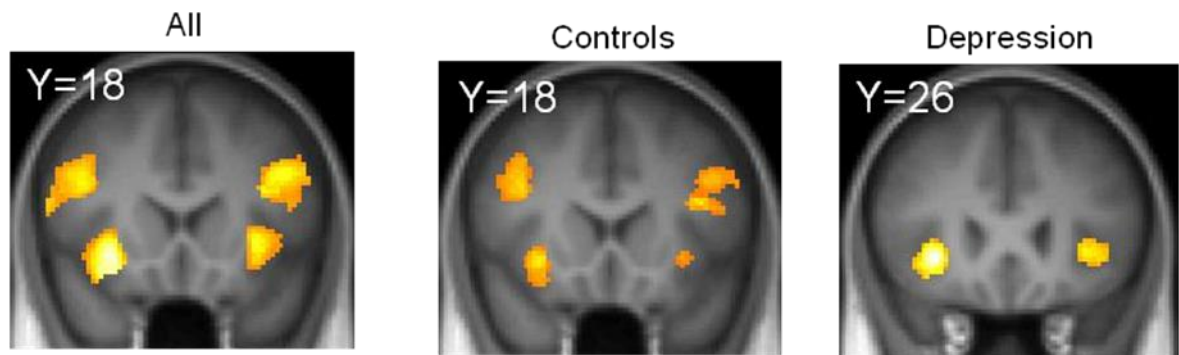
Supplementary figure S2

Neural responses for the contrast (CD>CC).



Supplementary figure S3

Neural responses for the contrast (DC>CC).



References

McClure, E. B., Parrish, J. M., Nelson, E. E., Easter, J., Thorne, J. F., Rilling, J. K., Ernst, M. & Pine, D. S. (2007). Responses to conflict and cooperation in adolescents with anxiety and mood disorders. *J Abnorm Child Psychol* **35**, 567-77.

Dr. Victoria Gradin
Center for Investigation of Basic Psychology (CIBPsi)
Faculty of Psychology
Universidad de la República
Montevideo 11200
(598) 24008555-285/286
victoriagradin@gmail.com

October 29, 2015

Dear Editors of Psychological Medicine

Thank you for the opportunity to re-submit our manuscript. Our detailed responses are below.

Reviewers' and editor's comments:

Reviewer #1: Methodologically competent and clearly written paper reporting specific contrasts in the Prisoner's Dilemma task, differing between depressed and well volunteers. The study is adequately powered and not confounded by medication. Important contribution to the behaviour and functional imaging literature of depressive disorder.

We thank the reviewer for these comments.

Reviewer #2: This paper reports findings from a Prisoner's Dilemma experiment in depression which is a potentially important contribution to the literature. Some important details about the experimental procedures are missing and emerging relevant literature is not always cited. There are also some concerns with the data analysis. Specific comments:

Introduction:

Some relevant recent literature is not cited. Page 5 ln 18: The assertion that stag hunt game is cognitively less demanding is problematic, especially in the light of two papers by Wako Yoshida et al., showing how demanding the computations of stag hunt interaction might be. If the authors want to raise this as a question, it would be important cover to both sides of the argument by citing these papers (2008 Plos Computational Biology; 2010 Journal of Neuroscience).

It has been reported (Emonds 2012 Social Neuroscience) that the Prisoner's Dilemma (PD) is associated with increased activation of the DLPFC when compared to the stag hunt game, possibly related to higher cognitive demands imposed by the PD. However, as suggested we have added in the same paragraph of the Introduction that the stag hunt game has been reported to imply demanding mental computations and the two references by Yoshida and colleagues have been added.

Methods

page 8 ln 25: Authors say that the participants "completed" Hamilton and MADRS. Please clarify whether the scales were undertaken by a rater or whether a self-report approach was used.

The Hamilton Depression scale and the Montgomery-Asberg Depression Rating were undertaken by a rater (one of the authors, VBG). This has been clarified in the text.

Page 9 ln 13: It would be a great benefit if the authors put the instructions for the experimental procedures as a part of the supplementary materials. For example, did they have a confederate design? What are the details of the manipulation, such that none of the participants suspected that they in fact interacted with a computerised partner? This would be very useful informative for future clinical studies, as well as allowing comparisons between relevant studies.

The specific instructions given to the participant on how to play the Prisoner's Dilemma have been added in the Supplementary Material. The participant was being told that he/she was playing with a co-player outside the scanner.

Page 9 ln 18: The authors say that people earned points in the game, whereas the experimental timeline in Fig 1 shows monetary amounts in £s. These should be consistent. It is also important to know whether the participants saw their accumulated earnings during the trials and whether or not they saw the opponent's winnings?

We thank the reviewer for asking this question which has allowed us to correct the text and figure. Participants were told that they would be accumulating earnings (pounds) during the game, although they were told that they would only be paid a percentage of these earnings. This has been clarified in the text.

While the initial screen design for the PD included displaying the £ symbol preceding the earnings in the pay off matrix, the £ symbol was finally not included in the version that was used in the scanner to keep the visual display as simple as possible. Figure 1 has been corrected and the £ symbol has been removed.

page 10 ln 7: It would be helpful to include the necessary information fully understand the computer algorithm, which drives the decisions in the experiment rather than simply referring to an earlier work. I suggest that authors design a flowchart of how the algorithm calculates probabilities through the trials and how these probabilities were drawn (eg by MATLAB?) and include this in supplementary materials. More detail about the probabilities and how they changed would be important to include

Information about the PD algorithm has been added in the Supplementary Material. The exact probabilities that the algorithm uses to compute a response on each trial are provided in this section.

Page 10 ln 17: The authors state that, the participants completed a questionnaire about their affective experiences to each of the Prisoner's Dilemma outcomes.

More detail about this questionnaire should be given, again perhaps in supplementary materials.

Detailed information about the questionnaire on emotional responses has been provided in the Supplementary Material.

Results

Page 13 ln 21: The authors have categorised each affective response and performed statistical analysis within each category independently of the analysis being performed on other categories. Unfortunately, this approach does not control for multiple comparisons so it may be preferable to use a Multivariate ANOVA approach: 5 categories of affective response (ie. satisfaction, betrayal, etc) x4 categories of outcomes (ie CC, CD etc) x2 clinical groups (control vs depression). This approach is robust for multiple comparisons and depending on the F statistics for main effects and interactions, the authors can progress with looking at the sub domains where interactions or main effects are observed.

As suggested, a three way ANOVA with factors emotion, outcome type and group has now been implemented and reported in the text. This analysis identified a significant effect of emotion ($F_{(2.18,93.77)} = 85.99, p < 0.001$), a significant effect of outcome ($F_{(2.5,107.50)} = 5.186, p = 0.004$), a significant emotion*group interaction ($F_{(2.18,93.77)} = 6.83, p = 0.001$) and a significant emotion*outcome interaction ($F_{(6.30,270.91)} = 65.18, p < 0.001$). No other main effect or interaction was found significant. The analyses that we previously had reported (independent ANOVAs for each emotion category) are now reported as follow up analyses.

page 14 [general]: Some of the degrees of freedom have decimal places, I think this implicitly suggest that the authors might be relying on Greenhouse-Geisser adjustment, but this is not explicitly stated or justified.

Where there was evidence for non-sphericity, the Greenhouse-Geisser correction was used. This has been noted in the text.

page 14 ln 18: I think it is quite valuable that the authors are reporting transition probabilities, which is the first time for a Prisoner's Dilemma study in depression. This deserves more emphasis.

Unfortunately we are constrained by maximum word limits plus we do not report between groups behavioural differences.

The authors frame their work in terms of cognitive control, but they make no mention of response time differences between the groups, which would give some further support for the argument.

We have now included analysis of reaction times. We examined reaction times for cooperation and defection following co-player cooperation or defection and having group as a factor. This analysis yielded no significant main effect for group or significant interactions with the group factor.

Of note, reaction times are not always reported on Prisoner's Dilemma studies and there is not a clear framework in how to interpret reaction times in the context of this task and particularly with regard to cognitive control. Therefore, we did not have specific *a priori* hypothesis regarding reaction times.

page 16 ln 14: The authors suggest a potential role of anger and betrayal, which would also require emotional regulation, yet only report correlations between DLPFC activity and guilt ratings. This is a somewhat selective way of reporting findings and doesn't feel right; the authors should also display the correlations between the DLPFC activity and anger and betrayal ratings separately, all side-by-side in the same figure.

We have now clarified in the text that there were no significant correlations between left DLPFC activity and anger and betrayal ratings for the CD outcome.

Discussion:

Again the authors should refer to recent relevant studies.

Two recent reviews have been added to the Discussion citations.

Wang, Y., Yang, L. Q., Li, S. & Zhou, Y. (2015). Game Theory Paradigm: A New Tool for Investigating Social Dysfunction in Major Depressive Disorders. *Front Psychiatry* 6, 128.

Pulcu, E. & Elliott, R. (2015). Neural origins of psychosocial functioning impairments in major depression. *Lancet Psychiatry* 2, 835-43.

Tables

Considering that the authors relied on alpha simulated cluster size thresholding, I think it is important to include cluster sizes in all of the neuroimaging tables and perhaps the tables could be sorted by the cluster size from the largest to the smallest.

We have included cluster sizes in all the neuroimaging tables. We believe though that it's of benefit for the reader to have the activations of interest in the first rows of the tables.

Supplementary table 1 in which the authors are reporting emotional and behavioural results should have a separate column for p values after suitable corrections.

We think this may be subject to misinterpretation and confusion. This is because the variables reported in the table have been analysed using ANOVAs with multiple factors. Therefore, it would be possible to report p-values related to the main effect of group, p values related to interactions including the group factor, or p-values related to follow up analysis after decomposition of interactions.

We have added a note in the Table advising the reader to refer to the main text for details on the statistical analysis of these variables.

Figures

Unfortunately, none of the bar charts in figure 1 have clearly visible y-axis and it is not possible to understand the values of the bars. Please provide a higher resolution figures in the resubmission.

The figures have been clarified.

Also, any decomposition of the F tests which reveals significant between group differences should be marked with an asterix on top of the error bars.

We think it may be confusing and potentially misleading to only highlight the significant between group differences that emerge after decomposition of the F tests, without also signalling the significant main effects and significant interactions, which is not really practical.

Reviewer #3: Please enter your comments to the ===AUTHOR=== here.

The authors used fMRI to investigate neural activity in unmedicated depressed participants (n=25) and well-matched healthy controls (n=25). During scanning, participants played the Prisoner's Dilemma. Depressed participants reported higher levels of negative feelings (betrayal, guilt) during the game than did controls. Neural activation compared between 'imbalanced' events (when one of the players cooperated and the other defected ('CD' and 'DC')) and 'draw' events (when both players either cooperated or defected ('CC' and 'DD')) found that depressed participants showed reduced activation in the left DLPFC, with the extent of signal reduction correlating with increased self-report feelings of guilt associated with DC outcomes.

This is a timely and original study, which has been well conducted, is clearly reported and has interesting results. I only have minor points of clarification to raise:

We thank the reviewer for these comments.

1. The fMRI data was modelled at both decision and outcome time points. What were these times?

The decision time corresponds to the time when the participant selected a column of the payoffs matrix. The outcome time corresponds to the time when the feedback screen with only one cell of the payoff matrix highlighted is presented. This has been clarified in the text.

2. Is the combined analysis of 'imbalanced' and draw' events standard or at least with some precedent in the literature?

Yes indeed, the contrast of imbalanced (CD and DC) versus draw (CC and DD) outcomes was analyzed in the Prisoner's Dilemma study by Rilling et al 2002 Neuron. This has been clarified in the text.

3. Given the correlation between DC outcomes and depression, was there a difference in activation during DC outcomes alone?

We report activations for the contrast (DC>CC). Across all participants this contrast showed significant activation in the bilateral DLPFC and bilateral insula (Supplementary Fig. S3, Supplementary Table S4), with no significant between-group differences observed in any regions of interest. This is reported in the Results section.

4. The patients also differed from controls in anxiety levels. It would be interesting to know if these correlated with activations, especially e.g. during imbalanced events.

There was no significant correlation in the depression group between left DLPFC activity and anxiety scores measured with the HAM-A or Spielberger State Anxiety scale.

Having made substantial changes to the text in accordance with recommendations, we hope it is now considered suitable for publication.

Yours sincerely

Dr Victoria Gradin